Automating Data Governance with Generative AI

Linus W. Dietz[1], **Arif Wider**[2], **Simon Harrer**[3,4]

[1]King's College London, [2]Hochschule für Technik und Wirtschaft Berlin, [3]innoQ Deutschland GmbH, [4]Entropy Data GmbH
linus.dietz@kcl.ac.uk, Arif.Wider@HTW-Berlin.de, simon.harrer@innoq.com

Abstract

The exchange of data across systems, both within and between organizations, is governed by company policies and data protection regulation. As policies and data flows evolve over time, ensuring continuous compliance of data exchange remains a complex challenge. In federated data architectures, the validation of data access requests is a critical and labor-intensive process. To formalize this task and enable automatic verification of compliance, rule-based constraint specification languages can be applied. However, data access constraints originate from legal documents, and translating them into formal data contracts is a tedious, repetitive, and error-prone task that can lead to inconsistencies and delays in maintaining compliance with the latest regulations. To address this challenge, we developed Governance AI, a large language model (LLM)-based tool for evaluating data access requests based on relevant policies, the type of data requested, and the request's context. To test our approach at scale, we propose an access request generator and a testing framework for computational data governance. In our evaluation, which involved 110 data access requests across two business domains, e-commerce and life insurance, we found that LLM-generated test cases were highly realistic, making them well-suited for comprehensive testing. Governance AI demonstrated a stricter approach than human experts, issuing a higher number of warnings and consistently flagging all critical cases where experts raised data sharing concerns. While the tool generated $3.6\times$more warnings than human experts, further inspection revealed that 80% of these warnings were classified as accurate. Our findings contribute to the automation of data governance by critically assessing the potential of generative AI in evaluating data access requests regarding legislation and internal policies.

Introduction

The increasing reliance on digital technologies has led to an increase in the collection and storage of sensitive personal data by organizations. While this data offers immense potential for insights and innovative applications, significant privacy challenges have also emerged, which led to increased regulation (European Parliament and Council of the European Union 2016; European Union 2024; California State Legislature 2018). The tension between leveraging data for utility and maintaining rigorous privacy protections has become a critical concern for organizations. Data sharing within complex organizational structures often requires navigating a delicate trade-off: sharing data too freely can compromise privacy, while overly restrictive sharing limits the potential for valuable analytics.

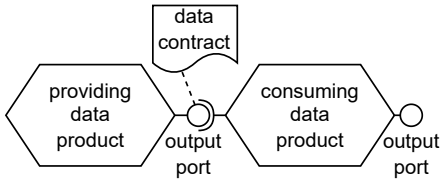


Figure 1: **Data products and their composition.** The sharing of data is defined through data contracts that specify the usage terms of the output ports.

To address this challenge, organizations must adopt effective data governance practices that enable secure and controlled data sharing. Instead of relying on costly and error-prone manual access management, data owners need automation tools that assist them in specifying correct access restrictions, allowing them to share data easily without compromising privacy. A modular data architecture, such as the *data mesh* architecture, facilitates the implementation of such automated governance support (Schultze and Wider 2021; Dehghani 2022; Wider, Jarmul, and Akhtar 2024).

Data mesh is a recent approach to enterprise data management that emphasizes domain-driven modularization, federated governance, and automation through self-service platforms (Dehghani 2020; Schultze and Wider 2021). The core unit of modularization in a data mesh is the *data product*. A data product is not merely a dataset managed with a product-focused mindset; it is also a data service designed for composability. Data products interact through output ports, which are clearly defined interfaces that specify the schema of the shared data. Any terms associated with the output ports, such as service-level agreements (SLAs), guarantees, or constraints, are outlined in corresponding *data contracts*. As illustrated in Figure 1, higher-order data products consume data from one or more source data products by connecting to their output ports, typically transforming the data internally, and potentially re-sharing their use-case-specific data through their own output ports.

Automation and verification of access restrictions, as well as consumers' adherence to sharing constraints, can be im-

plemented through clearly defined data contracts, whose constraints are verifiable. This can be achieved by using an open standard for the data contract structure, such as the Open Data Contract Standard (Bitol 2025). Conversely, data consumers also rely on the guarantees specified in the provider's data contract, including the structure, data quality, and SLAs (Wider, Jarmul, and Akhtar 2024).

Additionally, global privacy rules regarding the handling of personally identifiable information (PII), as defined through legislation or privacy policies, must be adhered to, ideally through automated verification. One way to automate this support is by describing contract-level constraints and global governance rules in a machine-readable format, such as using a formal constraint specification language. However, these constraints often originate from legal documents or internal policies. Translating these documents into formal data contracts is a tedious and error-prone process that must be repeated whenever the source documents are updated.

Leveraging the structure of the decentralized data mesh architecture and recent advances in generative AI, we propose a different approach to computational data governance. We use large language models (LLMs) to provide automated support for data governance tasks by directly processing legal documents, such as privacy policies, rather than requiring translation into formal contracts first. Thereby, we addressed the core challenge in data governance: determining whether data flows comply with the constraints defined in data contracts and global policies, i.e., the critical decision of whether data access should be granted. While we hypothesize that such a system should not fully replace human experts in this decision-making process, our goal was to reduce the workload of these experts by automating the verification of data access. Additionally, we aimed to assist them with warnings and suggestions regarding potential policy violations.

After reviewing the related work, we present three methodological contributions (C1-C3); the evaluation of which leads to interesting results (C4).

C1: We described the design of the Governance AI, an LLM-based tool that automatically checks access requests for policy violations in a data mesh architecture.

C2: We developed a testing framework for checking policy violations of data access requests at scale.

C3: We evaluate our Governance AI tool using a comprehensive set of $n = 110$ generated data access requests in two domains.

C4: Based on the results, we evaluate how the Governance AI tool compares to data governance experts in assessing access requests, as well as the accuracy of its warnings and suggestions.

Our findings have both theoretical and practical implications for the fast-growing field of computational data governance.

Related Work

Data mesh is a recent, industry-driven approach to decentralized enterprise data management (Dehghani 2022; Schultze and Wider 2021; Perrin and Broda 2024). Since data mesh originated in industry, research on the topic is still limited (Goedegebuure et al. 2024), with Machado et al. (Machado, Costa, and Santos 2021) being first to describe the conceptual framework of data mesh to the academic community. A main driver for adoption in industry is improved data governance efficiency (Bode et al. 2024; Oppold, Fritz, and Woltmann 2025), making data mesh a suitable conceptual framework for the goals of this article.

Computational Data Governance

An important concept of data mesh is *computational data governance*, which refers to the automation of governance tasks using tools provided by a mesh-wide data infrastructure platform (Dehghani 2020). For instance, if data product owners correctly tag sensitive data in their data products, the platform can automatically ensure adequate protections and policy compliance. Joshi et al. presented an industry case study on data governance in a data mesh setting but without much focus on automation (Joshi, Pratik, and Rao 2021). Wider et al. explored different approaches of how to automate governance tasks in data mesh platforms (Wider, Verma, and Akhtar 2023) and presented a concept for automated checking of data privacy constraints (Wider, Jarmul, and Akhtar 2024). By adopting the concept of sharer-driven contracts, i.e., data product owners define constraints for data consumers, we aim to automate the aspect of data governance that ensures data flows comply with policies and regulations.

The work presented by Dolhopolov et al. is similar to ours in the sense that it also implements an automated checking of data access attempts by enforcing governance policies (Dolhopolov, Castelltort, and Laurent 2024). However, in their work, these governance policies are formal rule-based policies and therefore first need to be created from the relevant legal policy documents. This introduces the problem of repeatedly translating legal documents into formal policies whenever the legal documents are updated. Furthermore, in their conceptual framework, Borovits et al. focus on data privacy within the context of computational governance (Borovits et al. 2023). While our work applies several of the concepts they presented, we improve upon this work by incorporating generative AI and providing a implementation and an evaluation of our approach.

An interesting alternative for managing personal data in a privacy-preserving way is AuthApp (Both et al. 2024), which leverages the principles of the SOLID data pods to enable GDPR-compliant data sharing (Dedecker et al. 2022; Sambra et al. 2016). Since this approach shifts control to the individual user, it represents a different direction compared to the enterprise-based approach described in our work.

Generative AI for Policy Compliance Testing

Generative AI, specifically LLMs appear to be well-suited for analyzing legal texts and making decisions based on given cases. In the seminal work of Bignotti and Camassa, the authors found that when GPT-4 was presented with historic Italian constitutional court rulings, the LLM was able to identify relevant articles and make consistent rulings with

minimal hallucinations (Bignotti and Camassa 2024). However, the authors also noted that the LLM exhibited a clear bias toward progressive interpretations of constitutional law. Herdel et al. (Herdel et al. 2024) used LLMs to generate AI application scenarios to test their compliance with the EU AI Act (European Union 2024). Similar to our test generation approach, generative AI has recently been applied to generate penetration tests to identify security issues. Hilario et al. highlight the creativity of the LLM in this context (Hilario et al. 2024). More broadly, generative AI has also been used to generate test cases for software testing (Shin 2024). Independent of the application of generative AI, our approach of feeding a list of data products to a large language model to generate test cases can also be seen as a form of model-driven testing (Schieferdecker 2012). Li and Maiti recently used an LLM for continuous compliance checks in the agricultural sector (Li and Maiti 2025), however, their work does not focus on privacy policies. Regarding privacy engineering, Amaral et al. used an LLM to check the completeness of a privacy policy with regard to standards set with the GDPR (Amaral et al. 2022).

To the best of our knowledge, our approach is the first to apply generative AI to support data governance in decentralized data architectures. Specifically, our method ensures that data flows within a data landscape comply with the relevant privacy policies.

Governance AI for Analyzing Data Access Requests

This section describes the design of the Governance AI, which is part of the commercial tool[1]. We first present the relevant concepts of the tool that are necessary for this work and then focus on how the Governance AI works in detail regarding automatically evaluating access requests for policy violations.

Basic Concepts

The commercial tool is an enterprise data marketplace with native support for data products and data contracts. At its core, it is designed to manage the metadata graph of a data mesh. The nodes are the data products with their output ports, and the edges are the access dependencies between data products. The meta model is illustrated in Figure 2.

Each data product is owned by a team and provides datasets through separate output ports, distinguished by technology, version, environment, and data model. The guarantees, such as data model specifications, usage constraints, and limitations, are defined within a data contract for each output port. The tool supports both the Open Data Contract Standard (Bitol 2025) and the Data Contract Specification (Christ and Harrer 2024) as a way to represent data contracts in YAML format.

An access request defines the dependencies within the data mesh graph. It represents a request by a consuming data product to access an output port of a providing data product. From the perspective of an access request, it differentiates between the consuming and providing data products, along with their respective owning teams. Members of the team responsible for the providing data product can grant or deny access to the consuming data product.

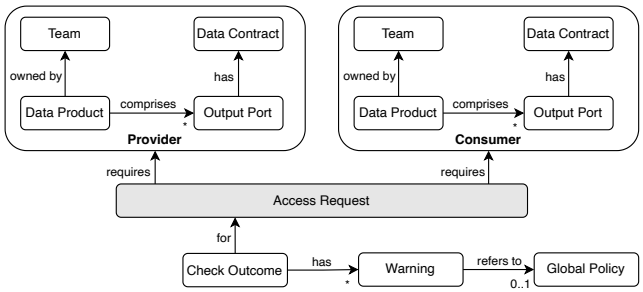


Figure 2: **Meta model of the elements involved in a data access request**. Each access request, if accepted, represents a data flow between a providing and a consuming data product. Governance AI examines the request for potential policy violations, producing a—potentially empty—list of warnings, each accompanied by a reference to the relevant section of the applicable global policy.

Global policies are rules that impose restrictions on the structure of the data mesh graph. These rules are defined in plain text and are not embedded within the graph itself. The tool automatically verifies whether data products, teams, data contracts, and access requests comply with these policies. In other words, it ensures that the entire graph adheres to the global policies. In Figure 2, we illustrate policy checks only for open access requests, as this is the primary focus of this paper. The outcome of these checks is a set of warnings that describe potential policy violations. Each warning either references a specific global policy or serves as a general safeguard, such as ensuring compliance with legal requirements. Additionally, a warning may include a suggested course of action to resolve the detected violation.

Detecting Policy Violations of Access Requests Using LLM-powered Automated Policy Checks

The Governance AI uses GPT-4o hosted on Microsoft Azure in Sweden with the default content filter. For each access request, the tool generates a set of warnings that highlight potential policy violations. This information is presented to the owner of the providing data product, who decides whether to approve or reject the request.

As stated in the introduction, a design principle is that a human has the last say. This is necessary for reasons of accountability, acceptance of AI solutions within organizations, and potential flaws of the employed technology. Figure 3 shows how the outcome of the Governance AI check is displayed to a data owner. The AI does not make decisions itself but instead generates warnings to assist decision-makers in making informed choices. Under the EU AI Act (European Union 2024), this system is categorized as "limited risk," meaning that users must be informed that the warnings are generated by an AI system.

The prompt consists of both the system prompt and the

[1]https://www.datamesh-manager.com

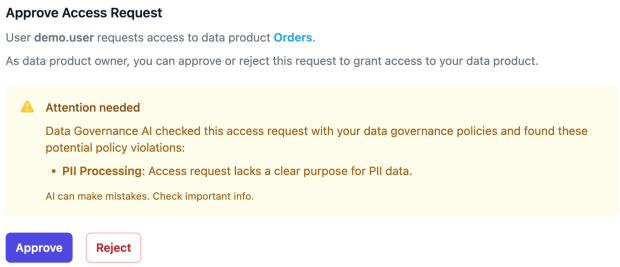


Figure 3: **Outcome of the Governance AI access request analysis.**

user prompt, which are outlined at a high level in Figure 4. The final version was developed through a systematic, iterative process in which each numbered component of the prompt (as shown in the figure) was individually modified and evaluated using a predefined test suite. We selected the configuration that maximized the F1-score for correctly accepting recommendations and warnings, while ensuring that no access requests deemed inappropriate by domain experts were accepted.

The system prompt provides general instructions about the task, the persona of the AI, and the steps to approach the task in a general manner. The full system prompt is shown in Figure 5. The user prompt contains all necessary metadata and context, including the access request, the providing and consuming data products, and the relevant global policies. Additionally, the user prompt includes more specific instructions for detecting policy violations and converting them into warnings with suggestions for remediation.

Whenever a data access request is submitted, it is automatically checked by the Governance AI. Additionally, the Governance AI checks can also be called through a dedicated API, which we used for testing purposes.

A Data Mesh Testing Framework for Automatically Checking Policy Violations

This section describes the design and implementation of an automated testing framework for data governance.

Design Considerations

In general, the testing framework's design followed a traditional approach to test automation by executing access request test cases and producing test results. Each test case is designed to run fully automated, in isolation from others, and the test results are intended to be reproducible. The latter property is practically unattainable due to the probabilistic nature of LLMs. The elements of an access request test case and its test result are as follows:

- Data Mesh
- Access Request
- Expected Outcome (No Objections, Warn)
- Actual Outcome (No Objections, List of Warnings, Error)

- **System prompt**
  1. **Task.** We describe the main task: analyzing access requests.
  2. **Persona.** We asked the model to adopt the persona of a Data Governance Expert.
  3. **Steps.** We describe the six steps how to analyze an access request.
- **User Prompt**
  1. **Access Request** that needs to be analyzed `(YAML)`
  2. **Provider** side of the access request, including the providing data product, the relevant output port, the data contract, and the providing team `(YAML)`
  3. **Consumer** side of the access request, including the consuming data product, all output ports, data contracts, and the consuming team `(YAML)`
  4. **Global Policies** governing the data mesh (`text`).
  5. **Detailed Instructions** about the task, the requirements, and additional constraints.
  6. **Required Elements** of the output with an explanation. The structure of the required elements was enforced using the "Structured Outputs JSON mode," cf. https://platform.openai.com/docs/guides/structured-outputs

Figure 4: **Structure of the access request analyze prompt.** The full prompt is provided in the supplementary material (https://github.com/LinusDietz/Automating-Data-Governance).

Your task is to analyze data access requests whether the consumer should be granted access to the data offered by the provider.

Respond as a data governance expert.

Follow these steps:
1. Understand the purpose of the access request of the consumer.
2. Understand the data contract of the producer.
3. Check that the purpose of the access request is in line with the data contract of the producer.
4. Understand the global policies of the organization.
5. Validate if this access request violates any policies.
6. Formulate warnings if and only if there is an obvious policy violation.

Make sure that the access request does not violate any policies or restrictions of the data contract of the producer.

Figure 5: **System Prompt of the Governance AI.**

Each test case describes a complete data mesh, which includes a specific access request to be checked, along with an expected outcome to be asserted. Conceptually, a data mesh can contain multiple access requests and can be reused across various test cases. The test execution flow for each test case has three steps, the "Setup Step," which resets the state to that of the data mesh of the test case, and the "Test Step," where the Governance AI runs the policy check for a specific access request. Finally, in the "Assert Step," the expected outcome is compared against the actual outcome.

Implementation

To make the test execution fully automated, we implemented the test framework in Python as a command-line application. The test framework requires a JSON file as an input that contains all the test cases. Each test case links to a folder containing all the YAML files describing the data products, data contracts, teams, and access requests of a particular data mesh, and it identifies a single access request via its ID to be checked within that data mesh. For each test case, the testing framework uses the tool's API to reset the state to that of the referenced data mesh, ensuring isolation between test case executions. It then calls the Governance AI API endpoint to analyze the access request and obtain a list of warnings in JSON format. An empty list of warnings means the actual outcome is "No Objections" and "Warn" otherwise. If the API call fails, e.g., due to errors in referencing data products, the actual outcome is an "Error". When all the test cases are executed, the test framework produces a JSON file containing all test cases together with their corresponding test result.

While LLMs have been shown to be relatively consistent in their output when handling complex cases (Bignotti and Camassa 2024), the probabilistic nature of LLMs prevents test execution from being strictly idempotent. As a result, the test framework does not provide a summary score for the executed tests. Instead, we created a dashboard that allows for quick visualization of each test case, highlighting whether the expected outcome matches the actual outcome.

Generating Systematic Access Request Test Cases

To put the testing framework into action, we collaborated with domain experts to create two *data maps* (i.e., sets of data products and their connections) in the insurance and e-commerce domains, allowing us to test realistic scenarios for evaluating our proposed system. These domains were chosen because they increasingly adopt data mesh architectures and the high importance of data governance (Ramos et al. 2024). In the subsequent step, we used these prototypical data maps to generate a comprehensive set of potential data access requests.

Creating and Validating Data Maps

The data maps were designed to establish a minimal yet complete Data Mesh architecture for a typical company in the insurance and e-commerce domains. Drawing from our experience, we began by creating central data products along with their respective provided data contracts. To ensure realism, we presented these data maps to a domain expert for validation.

To further enhance their authenticity, we incorporated real-world privacy policies from the experts' companies. Both experts agree that the publicly available privacy policies serve as the most relevant basis for determining whether data access can be granted to a different team within the organization. The reason is that whenever a contract is established between a company and a customer, the privacy policy becomes part of the agreement. Since the privacy policy defines the scope of data processing, e.g., requiring prior consent before contacting a user for marketing purposes, it imposes restrictions on internal data sharing, particularly for activities that require using PII. Notably, both policies were available to the system in the companies' primary language (German), while all other inputs, including access requests, were in English. In the following section, we briefly present the two data maps alongside the expert's design rationales for the insurance and e-commerce domains.

Insurance domain. In the insurance domain, our initial data map proposal underwent significant revisions by the domain expert, resulting in a final set of five providing data products: "Actor," "Benefits," "Life Insurance Contracts," "Marketing Campaigns," and "Underwriting Life Insurance." The expert explained that, due to legal constraints, different business branches within the same insurance group must remain separate, making sharing across company boundaries impermissible. As a result, they modeled the life insurance domain as it is a core business branch that has strict limitations with respect to data governance. Additionally, based on their experience, they noted that each domain or team typically maintains only one outward-facing data product. Each data product provides a single output port with a defined set of fields. For example, "Life Insurance Contracts" contains eight fields listed in Figure 1. For replication purposes, all data products are available in the data map directory of the supplementary material[2]. The domain expert holds the job title "Head of Data & AI Governance" at a large insurance firm and has 15 years of experience in the insurance industry.

Table 1: **Example of tagged fields in an output port.** Life Insurance Contracts from the insurance domain

Field	Type	Tags
policyholder name	`string`	PII
policyholder base data	`string`	PII
prior medical record	`string`	PII
sum insured	`string`	
premiums	`string`	
agent	`string`	
agent commission	`int`	trade secret
beneficiaries	`string`	PII

E-commerce domain. In the e-commerce domain, our

[2]https://github.com/LinusDietz/Automating-Data-Governance

- **System prompt**
  1. **Persona** We asked the model to adopt the persona of a Senior IT Governance Specialist in the respective domain
- **User Prompt**
  1. **Overview** of the task
  2. **Providing data product** including the data contract of the output port `(YAML)`
  3. **List of potential consuming teams.** This was to avoid generating access requests within the same team/domain. We provided the teams as `text` with their names and IDs.
  4. **Detailed Instructions** about the task and the requriements in plain `text`.
  5. **Privacy Policy** We specifically asked for a wide range of realistic access requests that can be accepted or should be rejected according to the privacy policy. We included the full privacy policy governing the data mesh into the prompt as unprocessed `text`.
  6. **Required Elements** of the output with an explanation. The structure of the required elements was enforced using the "Structured Outputs JSON mode"
  7. **Example** of a representative access request `(YAML)`

Figure 6: **Structure of the access request generation prompt.** Refer to Figure 2 for further details about elements in the data mesh. The full prompt is provided in the supplementary material (https://github.com/LinusDietz/Automating-Data-Governance) and will be published upon acceptance.

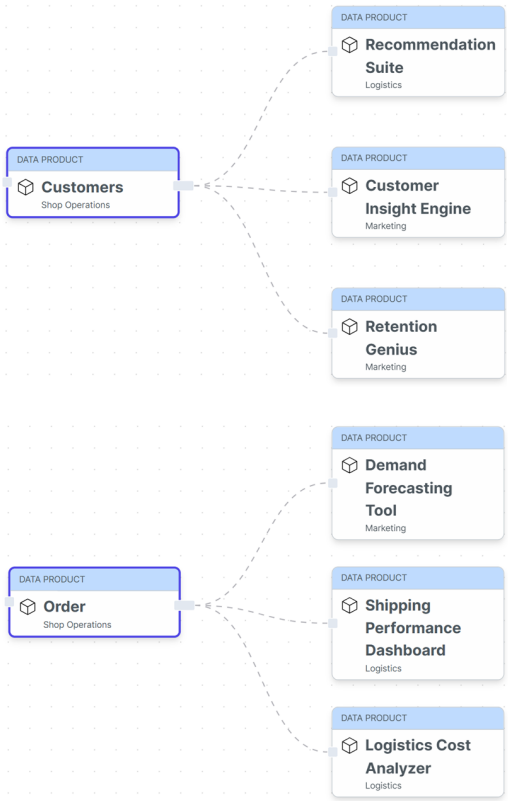


Figure 7: **Examples of access request scenarios generated for the e-commerce domain.** Overall, we generated 10 access requests for each providing data product.

expert proposed three teams, each responsible for multiple data products. *Shop Operations* provided the data products "Customers," "Order," "Order Lines," and "Payments"; *Logistics* managed "Products and "Fulfillment;" and *Marketing* was responsible for the (Marketing) "Campaigns" data product.

Similar to the insurance domain expert, they confirmed that this foundational data map was minimal yet representative and realistic for a typical e-commerce company. The e-commerce domain expert holds the job title "Head of Data Governance" at a large European e-commerce company and has nine years of experience in the industry.

Designing an LLM Prompt for Access Request Generation

Building on previous work on generating use cases with LLMs (Herdel et al. 2024), we aimed to design a prompt capable of producing a comprehensive list of access requests. Following an iterative development process informed by OpenAI's prompt engineering guidelines (OpenAI 2025), we informally evaluated the generated responses and experimented with different prompt configurations.

Ultimately, the prompt structure outlined in Figure 6 yielded satisfactory results, which we then evaluated with data governance experts from both domains. Based on our experience, providing the LLM with extensive context directly from the data mesh in YAML format proved to be important and made the generation of test cases fully automated. For example, our test suite generation script would iterate over all elements in the data map and dynamically input the relevant information, such as the fields in the output ports of the providing data products into the prompt. This property makes the prompt domain-agnostic, allowing it to be reused for generating access requests test suites for different data maps. Enforcing a structured output format using a JSON schema was essential to ensure complete information in all cases. The only element excluded from the final prompt was the entire data map, as its inclusion led to misaligned access requests. To ensure equal coverage within the data maps, we executed the prompt independently for each providing data product.

Creating Access Request Test Cases

One relevant design consideration in the access request generation step was that the output of the LLM is available in a machine-readable format that can automatically transformed into a YAML format needed to describe the access requests in the testing framework. Concretely, we used the LLM's response to create two YAML artifacts: A newly introduced

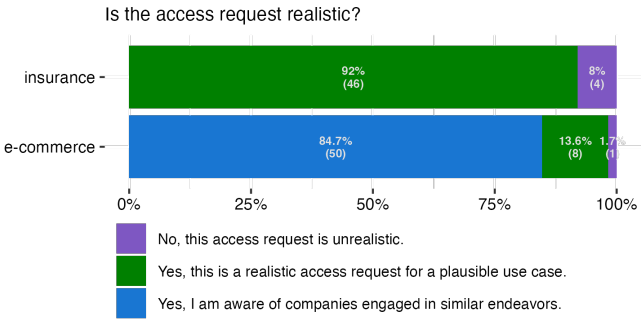


Figure 8: **Almost all LLM-generated access requests are realistic, with most already existing in the e-commerce domain.** The figure shows the distribution of the experts' assessment regarding the realism of access requests.

consuming data product representing a use case and a corresponding access request toward the providing data product. In Figure 7, we exemplify 6 such consuming data products for the "Customers" and the "Order" data products. Note that the consuming data products could also be data applications, however, in the scope of this work, we do not differentiate between data products and data applications. The dashed lines represent the pending access requests.

The LLM's machine-readable output enables automatic conversion into the YAML format required for testing the generated access requests. These can then be seamlessly added to the data map and executed within the testing framework.

Results

In total, we generated 10 use cases with respective access requests for each providing data product, resulting in 50 test cases in the insurance domain and 60 in the e-commerce domain. The number of test cases was limited by the available time of the domain experts to analyze them, yet with an overall 110 test cases, we ensured a comprehensive analysis. In the following sections, we evaluate them regarding realism, the Governance AI's ability to detect policy violations, and the usefulness of the Governance AI's warnings and suggestions.

Realism of Generated Access Requests

As the access requests and the underlying use cases were generated by an LLM, it is important to validate that the access requests are realistic and the underlying use cases are plausible in the respective domains. For this reason, in the first step, we asked the experts to assess the access requests and whether they are realistic to eliminate any potentially hallucinated test cases. We asked the experts to classify them as either already existing, i.e., they are aware companies engaged in similar activities, generally realistic ones, and finally, unrealistic ones, which we would discard for the subsequent steps.

In both domains, the vast majority of access requests were classified by the experts as realistic, with 85% already existing in the e-commerce domain (Figure 8). In the e-commerce domain, only one access request was discarded, whereas in the insurance domain, it was 4. Analyzing the experts' rationale for excluding use cases, two were ethically dubious or incompatible with law, e.g., the EU AI Act (European Union 2024), one described nonsensical marketing efforts, and two were practically infeasible operational improvements. The expert in the insurance domain did not differentiate between existing and realistic use cases, due to a misunderstanding of the distinction. However, we ensured that the inclusion/exclusion criteria were correctly understood.

Takeaway: Almost all LLM-generated access requests are realistic, confirming the merit of LLM-based generation of the test cases.

Detection of Policy Violations

Using the previously introduced testing framework, we systematically evaluated all realistic access requests regarding policy violations using the Governance AI.

The output of the Governance AI was compared against the expert baseline, resulting in four possible outcomes. In the first two cases, the expert and the system agreed: either determining that there were "no objections" to the access request or identifying issues that warranted a "warning." In cases of disagreement, the Governance AI either issued a warning while the expert had no objections or vice versa. In Figure 9 the four scenarios are summarized by means of confusion matrices, with e-commerce being on the left side and the insurance domain on the right. The numbers in the fields represent the absolute and relative counts.

Note that there are no cases in the bottom-left "missed warnings" quadrant, where the expert had objections, but the Governance AI did not formulate any.

In the **e-commerce domain**, the expert argued that all access requests are reasonable and would realistically be accepted without objections. However, they mentioned that this assessment is based on the assumption that only such personal data is available for analytics for which the user has given their consent to marketing purposes and exchange with third-parties. This is because in the company of the expert, data is collected centrally and only made available for analytics purposes after having been filtered for appropriate user consent. This is a common practice in organizations with a centralized data architecture. However, in a decentralized data mesh setting, data is shared in a more federated way, and the data owner has to ensure that the data is shared in a privacy-preserving way so that such a global assumption can usually not be made. The Governance AI is optimized for this decentralized setting and therefore assessed the access requests without this consent-given assumption. Nevertheless, the Governance AI agreed with the expert's take in 74.1% of the cases, however, formulated warnings in the remaining 15 access requests.

In the **insurance domain**, the expert formulated warnings for 12 (26%) of the access requests and accepted the other 74%. Again, the Governance AI was stricter, producing warnings in 41 of 46 access requests, 29 of which the expert found to be unproblematic.

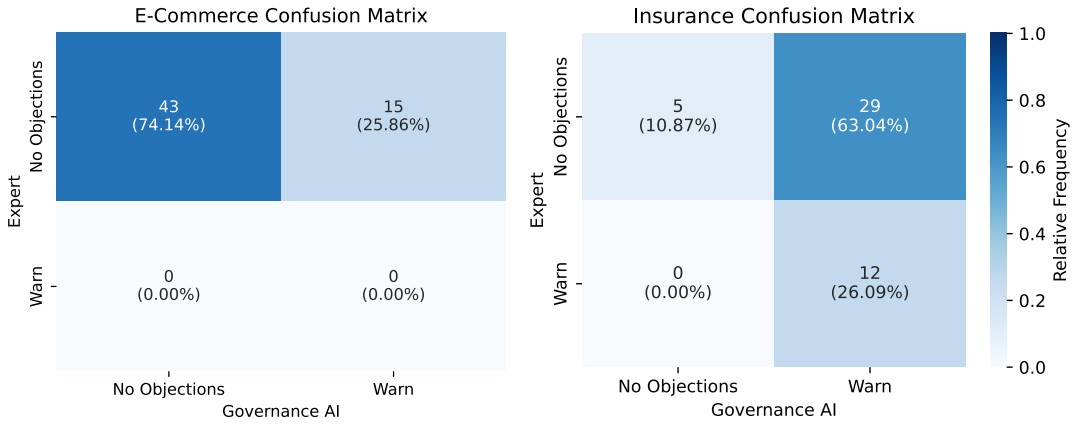


Figure 9: **The confusion matrices for the two domains show different patterns of agreement.** Crucially, the Governance AI never failed to issue a warning in cases where the expert deemed one necessary, however, it produced 3.6 times more warnings than the human expert.

Takeaway: The Governance AI is generally stricter than the expert, leading to a higher number of warnings. Crucially, the Governance AI never fails to issue a warning in cases where the expert deems one necessary.

Usefulness of Governance AI Warnings and Suggestions

To better understand the differences in judgment between the experts and the Governance AI, we presented all access requests with warnings to the experts again, this time alongside the warnings themselves and the AI's suggested resolution.

Interestingly, in both domains, 80% of the warnings were labeled as correct. However, only a minority, 25% in the insurance domain and 33% in the e-commerce domain, were deemed entirely correct, while the rest contained some flaws in their reasoning. The experts classified the remaining 20% of warnings as incorrect, further dividing them into two categories: generally valid objections that were nonetheless incorrect in the specific context of the access request (8–13%) and warnings that addressed non-existent issues or were entirely out of context. The distribution is visualized in Figure 10(a).

The e-commerce domain expert commented on the warnings that *"[the AI] follows a GDPR-centric approach that always assumes the "worst-case scenario." While this conservative stance is fundamentally not incorrect, the AI will consistently act as a perpetual naysayer."* We believe this a valuable property of the Governance AI, as GDPR violations can be really expensive and might even lead to bankruptcy.

The Governance AI also provided brief suggestions on how access requests could be improved for every warning. The distribution of response as shown in Figure 10(b) highlights the inherent differences between the two domains. In the e-commerce domain, issues were relatively easy to resolve by, for example, limiting data access to specific fields or anonymizing data before processing. In contrast, the suggestions in the insurance domain were rated as less useful. Specifically, while 53% of the suggestions were considered correct, 43% were deemed misleading since these would not effectively resolve the underlying issue. The insurance domain expert commented on this as follows: *"The warnings and suggestions are superficially acceptable but insufficient in detail and from a legal perspective."* What both domains have in common, however, is that most suggestions state that the warnings can be fixed by using anonymized non-PII instead of PII data.

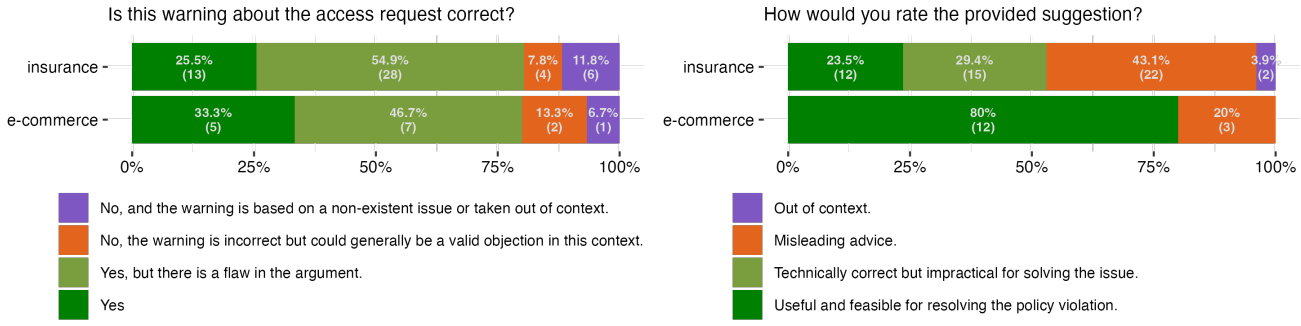
Takeaway: The warnings issued by the Governance AI were largely accurate, however, in a domain where privacy is highly relevant, it is hard to resolve policy violations in an easy way.

Discussion

Theoretical Implications

Since the majority of test cases were deemed realistic by domain experts, it is reasonable to assume that our test case generation process could produce an even higher number of realistic cases until reaching a saturation point. To ensure a thorough evaluation by domain experts, we limited our analysis to 10 test cases per providing data product. This finding aligns with previous research, which showed that 70% of LLM-generated use cases for facial recognition and analysis were already in existence, while 30% were classified as "upcoming" (Herdel et al. 2024).

Calibrating a warning system depends heavily on its real-world deployment, particularly how users interact with it (Fu et al. 2020; Lee and See 2004). Excessive warnings can lead to users ignoring them (Breznitz 1984), while insufficient sensitivity may result in failing to flag critical issues, leading to adverse consequences (Hall 2000). In this study, false warnings had relatively minor consequences, as the final decision to approve a data-sharing request remained with the data owner, who could choose to disregard the warning. Notably, our results indicated that in 26% of test cases in the e-commerce domain and 63% in the insurance domain, experts initially had no objections to access requests that the



(a) Expert assessment of the warnings

(b) Expert assessment of the suggestions

Figure 10: **The warnings issued by the Governance AI were largely accurate.** However, the more stringent data protection requirements in the insurance domain made it challenging to generate meaningful suggestions for resolving critical data access requests—whereas this was more feasible in the e-commerce domain.

Governance AI flagged with warnings (top-right quadrant of the confusion matrices in Figure 9). However, upon further review of the warnings issued for access requests they had initially not objected to, both experts agreed at a similar rate of 80% that the warnings were justified (Figure 10(a)).

On the other hand, failing to detect policy violations can lead to privacy breaches, with fines that vary based on severity and jurisdiction (European Parliament and Council of the European Union 2016). Notably, our findings show that the Governance AI never failed to issue a warning in cases where the expert deemed one necessary (bottom-left quadrant of the confusion matrices in Figure 9). This means that the Governance AI was always at least as "strict" as the domain expert. Any attempts to calibrate the AI's warning threshold to increase the true "no objections" rate, which in our study was 74% in the e-commerce domain but only 11% in the insurance domain, must ensure that this property is preserved to prevent overreliance on the assistive system (Alberdi et al. 2009; Inagaki and Itoh 2013).

Practical Implications

This study presented an approach to automatically ensuring policy compliance using generative AI by analyzing data access requests in a federated data mesh. We advanced the state of the art by proposing a system that automatically checks access requests. The Governance AI produced more warnings than an expert would find, especially in the domain, where the protection of PII or other sensitive data is critical.

The fact that Governance AI was at least as strict as human experts could be practically leveraged to introduce a higher level of automation, i.e., automatically approving access requests when Governance AI does not produce any warnings. While our results support this potential transfer of decision-making to the system, we caution against adopting this idea without careful consideration. A conclusive assessment would likely require further studies and additional safeguards (Sheridan 2002; Lee and See 2004).

Interestingly, the Governance AI always provided a suggestion for every warning, even if there was nothing that could be done to rectify the access request. It is similar to a person who, when asked for directions, offers an answer, even if they don't know the way. This behavior is an inherent property of current LLMs, and while it is an area of ongoing research (Zhang et al. 2024), we could not address it by simply rewriting the Governance AI prompt. Instead, it may be more effective to introduce an additional LLM step to qualify the suggestion, potentially removing it if categorized as inappropriate.

The automation of the tooling around the Governance AI opens up additional possibilities. It could be leveraged to determine the implications of proposed changes in the policy text. That way, the Governance AI opens a fast feedback loop to the legal and governance experts drafting those policies before the policy changes are put into effect, eventually leading to better policies. Moreover, the Governance AI opens the possibility to continuously monitor evolving data meshes for arising policy violations. This allows otherwise probably undetected policy violations caused by adding data sharing restrictions for already approved data flow to be detected shortly after the additional restrictions are defined.

Finally, based on the suggestions from both Governance AI and domain experts, it should become standard practice for data products handling PII data to offer a secondary, less sensitive output port containing anonymized or aggregated data. This alternative output port would allow consumers to access data that meets their needs without compromising privacy, particularly for use cases that can be fulfilled with non-sensitive data.

Limitations & Future Work

We evaluated 46 and 58 realistic use cases across two domains where data mesh architectures are increasingly being introduced. While these test cases represent a considerable sample and cover many business processes within the two model companies, our findings are currently limited to these two domains. We observed that data governance in different sectors involves distinct utility-risk trade-offs, with e-commerce appearing to offer more flexibility compared to the more restricted life insurance industry. Therefore, it would be interesting to expand this study to further domains,

such as business-to-business domains, and also domains that have been impacted by increasing privacy regulation, such as finance, and healthcare (Layton and Elaluf-Calderwood 2019). Due to the limitations of obtaining detailed information about real-world company data architectures, we used the data products of a model company's data mesh as the basis for our testing. This had the advantage, that the results are more generalizable to other companies, i.e., we could sidestep company-specific exceptions in how they handle data. However, in the future, we plan to establish data governance auditing methods based on the contributions of Nabar et al. (2008). For this purpose, it would be interesting to develop test coverage metrics for the data contracts and privacy policies. This would also give more informed indications about the "sufficient" number of test cases.

Analyzing the workflows of data product owners, requesters, or legal teams in formulating, assessing, and deciding on data access requests was beyond the scope of this paper. Our analysis focused on the realism of access requests rather than how well they were formulated. In practice, we know that different authors formulate access requests in differenty ways. An opportunity for future research would be the quantification of the benefits and drawbacks of AI-based computational governance in terms of human aspects, business needs, efficiency, and acceptance by its users (Allaham, Kieslich, and Diakopoulos 2025).

In this context, further research is needed to explore how users with varying levels of expertise interact with the GovernanceAI system. This is particularly important given the potential sociotechnical risks, such as user overreliance and long-term habituation, which may lead to unintended consequences over time that need to be understood.

Conclusions

In this paper, we proposed Governance AI, an LLM-based approach to automating data governance by evaluating data access requests against privacy policies and data contracts within a decentralized data mesh architecture. Our study demonstrated that the system effectively identified policy violations, issuing more warnings than human experts while never failing to flag critical cases. Although Governance AI applied stricter assessments, the expert review confirmed that 80% of its warnings were valid. Thus, we demonstrated its potential to assist data governance experts in assessing a large number of LLM-generated, yet realistic, data access requests.

Governance AI could serve as a central component of a data governance management tool capable of monitoring and managing an entire data mesh. The system was designed to consider all relevant policies without preprocessing steps and to track changes within the data ecosystem to prevent data breaches proactively. By leveraging the testing framework, operators of decentralized data architectures, such as multinational corporations with varying privacy policies across markets, could evaluate the implications of different policies through comprehensive test suites stemming from the access request generation tool.

Additionally, Governance AI could function as an educational tool, guiding users in drafting data access requests that comply with all applicable regulations. It could provide suggestions for formulating access requests and use cases that adhere to regulatory requirements or recommend alternative data products containing less sensitive information.

Despite our study's encouraging results, implementing computational data governance remains a complex challenge (Perrin and Broda 2024; Dolhopolov, Castelltort, and Laurent 2024). The ability to handle multilingual natural language underscored the potential of generative AI in data governance. Our Governance AI tool and testing framework represent a step toward more effective implementations of computational governance in modern enterprise data architectures.

Ethical Considerations Statement

This work explores the use of GenAI, specifically LLMs, to assist with automated data governance in enterprise environments. While our system, Governance AI (classified as "limited risk" under the EU AI Act (European Union 2024)), does not make final decisions, it provides policy violation warnings that could influence decisions regarding data access. To mitigate ethical risks, human oversight is embedded as a design principle, ensuring that accountability and final decision-making authority remain with human decision-makers.

The transfer of information for the purposes of checking access requests is governed by the individual contracts between users of https://www.datamesh-manager.com as well as the terms of services the platform has with LLM providers.

References

Alberdi, E.; Strigini, L.; Povyakalo, A. A.; and Ayton, P. 2009. Why Are People's Decisions Sometimes Worse with Computer Support? In Buth, B.; Rabe, G.; and Seyfarth, T., eds., *Computer Safety, Reliability, and Security*, 18–31. Berlin, Heidelberg: Springer. ISBN 978-3-642-04468-7.

Allaham, M.; Kieslich, K.; and Diakopoulos, N. 2025. Global Perspectives of AI Risks and Harms: Analyzing the Negative Impacts of AI Technologies as Prioritized by News Media. arXiv:2501.14040.

Amaral, O.; Abualhaija, S.; Torre, D.; Sabetzadeh, M.; and Briand, L. C. 2022. AI-Enabled Automation for Completeness Checking of Privacy Policies. *IEEE Transactions on Software Engineering*, 48(11): 4647–4674.

Bignotti, C.; and Camassa, C. 2024. Legal Minds, Algorithmic Decisions: How LLMs Apply Constitutional Principles in Complex Scenarios. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 7(1): 120–130.

Bitol. 2025. *Open Data Contract Standard (ODCS)*. LF AI & Data Foundation. Version 3.0.1.

Bode, J.; Kühl, N.; Kreuzberger, D.; and Holtmann, C. 2024. Toward Avoiding the Data Mess: Industry Insights From Data Mesh Implementations. *IEEE Access*, 12: 95402–95416.

Borovits, N.; Kumara, I.; Tamburri, D. A.; and Van Den Heuvel, W.-J. 2023. Privacy Engineering in the

Data Mesh: Towards a Decentralized Data Privacy Governance Framework. In *International Conference on Service-Oriented Computing*, 265–276. Springer.

Both, A.; Kastner, T.; Yeboah, D.; Braun, C.; Schraudner, D.; Schmid, S.; Käfer, T.; and Harth, A. 2024. AuthApp – Portable, Reusable Solid App for GDPR-Compliant Access Granting. In Stefanidis, K.; Systä, K.; Matera, M.; Heil, S.; Kondylakis, H.; and Quintarelli, E., eds., *Web Engineering*, 199–214. Cham: Springer. ISBN 978-3-031-62362-2.

Breznitz, S. 1984. *Cry wolf: The psychology of false alarms*. Psychology Press.

California State Legislature. 2018. California Consumer Privacy Act of 2018.

Christ, J.; and Harrer, S. 2024. Data Contract Specification. https://datacontract.com. [Online; accessed 10-March-2025.

Dedecker, R.; Slabbinck, W.; Wright, J.; Hochstenbach, P.; Colpaert, P.; and Verborgh, R. 2022. What's in a Pod? – A knowledge graph interpretation for the Solid ecosystem. In Saleem, M.; and Ngonga Ngomo, A.-C., eds., *6th Workshop on Storing, Querying and Benchmarking Knowledge Graphs*, volume 3279 of *CEUR Workshop Proceedings*, 81–96.

Dehghani, Z. 2020. Data mesh principles and logical architecture. *MartinFowler.com*.

Dehghani, Z. 2022. *Data Mesh: Delivering Data-Driven Value at Scale*. O'Reilly Media, Inc. ISBN: 9781492092391.

Dolhopolov, A.; Castelltort, A.; and Laurent, A. 2024. Implementing Federated Governance in Data Mesh Architecture. *Future Internet*, 16(4).

European Parliament; and Council of the European Union. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council.

European Union. 2024. Artificial Intelligence Act (AI Act): Regulation (EU) .../... of the European Parliament and of the Council laying down harmonized rules on artificial intelligence and amending certain Union legislative acts. Official Journal of the European Union (OJ). Adopted version, 2024.

Fu, E.; Johns, M.; Hyde, D. A. B.; Sibi, S.; Fischer, M.; and Sirkin, D. 2020. Is Too Much System Caution Counterproductive? Effects of Varying Sensitivity and Automation Levels in Vehicle Collision Avoidance Systems. In *CHI Conference on Human Factors in Computing Systems*, CHI'20, 1–13. New York, NY, USA: ACM. ISBN 9781450367080.

Goedegebuure, A.; Kumara, I.; Driessen, S.; Van Den Heuvel, W.-J.; Monsieur, G.; Tamburri, D. A.; and Nucci, D. D. 2024. Data Mesh: A Systematic Gray Literature Review. *ACM Computing Surveys*, 57(1).

Hall, S. 2000. Psychological consequences for parents of false negative results on prenatal screening for Down's syndrome: retrospective interview study. *BMJ*, 320(7232): 407–412.

Herdel, V.; Šćepanović, S.; Bogucka, E.; and Quercia, D. 2024. ExploreGen: Large Language Models for Envisioning the Uses and Risks of AI Technologies. *AAAI/ACM Conference on AI, Ethics, and Society*, 7: 584–596.

Hilario, E.; Azam, S.; Sundaram, J.; Imran Mohammed, K.; and Shanmugam, B. 2024. Generative AI for pentesting: the good, the bad, the ugly. *International Journal of Information Security*, 23(3): 2075–2097.

Inagaki, T.; and Itoh, M. 2013. Human's Overtrust in and Overreliance on Advanced Driver Assistance Systems: A Theoretical Framework. *International Journal of Vehicular Technology*, 2013: 1–8.

Joshi, D.; Pratik, S.; and Rao, M. P. 2021. Data governance in data mesh infrastructures: the Saxo Bank case study. In *21st International Conference on Electronic Business (ICEB)*, 599–604.

Layton, R.; and Elaluf-Calderwood, S. 2019. A Social Economic Analysis of the Impact of GDPR on Security and Privacy Practices. In *12th CMI Conference on Cybersecurity and Privacy*, 1–6. IEEE.

Lee, J. D.; and See, K. A. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46(1): 50–80.

Li, J.; and Maiti, A. 2025. Applying Large Language Model Analysis and Backend Web Services in Regulatory Technologies for Continuous Compliance Checks. *Future Internet*, 17(3): 100.

Machado, I. A.; Costa, C.; and Santos, M. Y. 2021. Data Mesh: Concepts and Principles of a Paradigm Shift in Data Architectures. In *International Conference on Enterprise Information Systems, Braga, Portugal*, volume 196 of *Procedia Computer Science*, 263–271. Elsevier.

Nabar, S. U.; Kenthapadi, K.; Mishra, N.; and Motwani, R. 2008. *A Survey of Query Auditing Techniques for Data Privacy*, 415–431. Boston, MA: Springer. ISBN 978-0-387-70992-5.

OpenAI. 2025. Prompt Engineering – Enhance results with prompt engineering strategies. https://platform.openai.com/docs/guides/prompt-engineering. [Online; accessed 6-March-2025.

Oppold, S.; Fritz, M.; and Woltmann, L. 2025. Data Contracts to Leverage (De-)centralized Data Management in Manufacturing Industries: An Experience Report. In *Datenbanksysteme für Business, Technologie und Web*, BTW'25, 731–743. Gesellschaft für Informatik.

Perrin, J.-G.; and Broda, E. 2024. *Implementing Data Mesh*. O'Reilly Media. ISBN 9781098156220.

Ramos, I.; Santos, M. Y.; Joshi, D.; and Pratik, S. 2024. Data Mesh Adoption: A Multi-case and Multi-method Readiness Approach. In Papadaki, M.; Themistocleous, M.; Al Marri, K.; and Al Zarouni, M., eds., *Information Systems*, 16–29. Cham: Springer. ISBN 978-3-031-56481-9.

Sambra, A. V.; Mansour, E.; Hawke, S.; Zereba, M.; Greco, N.; Ghanem, A.; Zagidulin, D.; Aboulnaga, A.; and Berners-Lee, T. 2016. Solid: a platform for decentralized social applications based on linked data. Technical report, MIT CSAIL & Qatar Computing Research Institute.

Schieferdecker, I. 2012. Model-Based Testing. *IEEE Software*, 29(1): 14–18.

Schultze, M.; and Wider, A. 2021. *Data Mesh in Practice*. O'Reilly Media, Inc. ISBN: 9781098108496.

Sheridan, T. B. 2002. *Humans and Automation: System Design and Research Issues*. USA: Wiley. ISBN 0471234281.

Shin, W. 2024. Utilizing Generative AI for Test Case Generation: Comparative Analysis and Guidelines. *International journal of advanced smart convergence*, 13(4): 145–154.

Wider, A.; Jarmul, K.; and Akhtar, A. 2024. Towards Automating Federated Data Governance. In *2024 IEEE International Conference on Web Services (ICWS)*, 10–19. IEEE.

Wider, A.; Verma, S.; and Akhtar, A. 2023. Decentralized Data Governance as Part of a Data Mesh Platform: Concepts and Approaches. In *2023 IEEE International Conference on Web Services (ICWS)*, 746–754. IEEE.

Zhang, H.; Diao, S.; Lin, Y.; Fung, Y.; Lian, Q.; Wang, X.; Chen, Y.; Ji, H.; and Zhang, T. 2024. R-Tuning: Instructing Large Language Models to Say 'I Don't Know'. In Duh, K.; Gomez, H.; and Bethard, S., eds., *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 7113–7139. Mexico City, Mexico: ACL.